Critical Review

An Adaptive Environmental Effects Monitoring Framework for Assessing the Influences of Liquid Effluents on Benthos, Water, and Sediments in Aquatic Receiving Environments

Keith M Somers,†‡ Bruce W Kilgour,† Kelly R Munkittrick,§|| and Tim J Arciszewski#*

†Kilgour & Associates, Ottawa, Ontario, Canada
‡Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada
§Canada's Oil Sands Innovation Alliance, Calgary, Alberta, Canada
||Present address: Wilfrid Laurier University, Waterloo, Ontario, Canada
#Alberta Energy Regulator (AER), Calgary, Alberta, Canada

ABSTRACT

Environmental effects monitoring (EEM) has been traditionally used to evaluate the effects of existing facilities discharging liquid effluents into natural receiving waters in Canada. EEM also has the potential to provide feedback to an ongoing project in an adaptive management context and can inform the design of future projects. EEM, consequently, can and should also be used to test the predictions of effects related to new projects. Despite EEM's potential for widespread applicability, challenges related to the effective implementation of EEM include the use of appropriate study designs and the adoption of tiers for increasing or decreasing monitoring intensity. Herein we describe a template for designing and implementing a 6-tiered EEM program that utilizes information from the project-planning and predevelopment baseline data collection stages to build on forecasts from the initial environmental impact assessment project-design stage and that feeds into an adaptive management process. Movement between the 6 EEM tiers is based on the exceedance of baseline monitoring triggers, forecast triggers, and management triggers at various stages in the EEM process. To distinguish these types of triggers, we review the historical development of numeric and narrative triggers as applied to chemical (water and sediment) and biological (plankton, benthos, fish) endpoints. We also provide an overview of historical study design issues and discuss how the 6 EEM tiers and associated triggers influence the temporal-spatial experimental design options and how the information gained through EEM could be used in an adaptive management context. *Integr Environ Assess Manag* 2018;14:552–566. © 2018 SETAC

Keywords: Environmental effects monitoring Environmental impact assessment Triggers Adaptive management Adaptive monitoring

INTRODUCTION

In the early 1990s in Canada, the Environmental Effects Monitoring (EEM) Program was initiated by the federal government under the Fisheries Act to evaluate the influences of liquid effluents (from pulp mills in 1992 and metal mines in 2002) on receiving aquatic environments (Walker et al. 2003). Canadian EEM programs principally focused on the assessment of fish populations and benthic invertebrates (i.e., benthos was used as a surrogate measure of the quality of fish habitat, e.g., sediment) through the comparison of samples of fish and benthos from areas "exposed" to effluent with samples from unexposed reference or "control" areas (Glozier et al. 2002). EEM programs were and are focused on the detection of biological effects (i.e., effects based, per Dubé and Munkittrick 2001) rather than the detection of chemical change (i.e., stressor based). These Canadian programs are somewhat akin to retrospective risk assessment (RRA; Suter 1993), wherein the ecological effects of exposure to a stressor are evaluated through combinations of field study and laboratory experimentation (see also Dubé and Munkittrick 2001; den Besten et al. 2003). The EEM approach is also akin to effectiveness monitoring (EM), which is the term often applied to ecological studies designed to determine if restoration efforts have succeeded (e.g., Chapman and Underwood 2000; Block et al. 2001). Consequently EEM-type analyses, as a catch-all term for both RRA and EM, have been used for assessing the effects of municipal wastewaters (Kilgour et al. 2005), oil and gas extraction (Parsons et al. 2010; DeBlois et al. 2014), deforestation (Wissmar 1993; Kreutzweiser et al. 2005), urbanization (Scarsbrook et al. 2000; Collier et al. 2009), hydroelectric facility operation (Richter et al. 1996; White et al. 2011), and other pressures on aquatic environments (Green 1979; Cairns et al. 1993).

EEM has been traditionally used to evaluate the effects of existing facilities that discharge liquid effluents into natural

receiving waters (Walker et al. 2003). When implemented appropriately, EEM is a powerful approach and can (and should) be used to test the predictions made in environmental impact assessments (EIA) related to new projects (Duinker 1989; Buckley 1991; Smith 1991; Dipper et al. 1998; Wood et al. 2000). However, the integration of EEM with EIA is currently not common (Arciszewski et al. 2017). Consequently, EEM has the potential to provide feedback to a project in an adaptive management context and to inform and improve future EIAs for similar and related projects.

Despite the fact that EEM programs are routinely implemented, agencies (and proponents of projects and EEM programs) routinely disagree on what will be measured, where and when measurements will be made, how the data should be summarized, and how to define an "effect" (be it "large" or "small"). Environment Canada (2012), for example, provides guidance on study designs, including the recommended number of sampling areas required in reference and exposure conditions and defined critical effect sizes. The minimum acceptable design includes 1 reference and 1 exposure area, despite decades of peer-reviewed literature urging the use of more complex designs that more robustly describe reference-area variability (Green 1979; Stewart-Oaten et al. 1986; Underwood 1991, 1994; Underwood and Chapman 2003). Environment Canada's (2012) guidance, based on metal-mining effluent regulation (MMER), further requires that assessments be based on only the current year (i.e., exposure period) to the exclusion of data from the baseline period or prior exposure periods (but see Jensen 1973). Huebert et al. (2011) and Mackey et al. (2012) identified that the minimum design has a good probability of resulting in the detection of differences that are unrelated to exposure to effluent and repeated the need for better design and the use and interpretation of data from more comprehensive study designs.

EEM-style monitoring in Canada's oil sands region has been even more contentious (Kelly et al. 2009, 2010; Schindler 2010). The Regional Aquatics Monitoring Program (RAMP) was an industry response to a provincial requirement for proponents to monitor operations and potential aquatic environmental effects at a regional scale (Hatfield et al. 2009). RAMP was essentially an EEM program designed collaboratively by the industry and representatives from provincial and federal government agencies. Monitoring (i.e., the selection of endpoints, sampling locations, sampling frequencies, and data analysis procedures) was generally designed with the intent to test hypothesized effects of oil sands operations on aquatic environments. Monitoring of the aquatic environment consisted of surveys of water, sediment, and biota (i.e., benthos, fish communities, and sentinel fish populations) and was implemented for 15 years before an external peer review in 2010 (Oil Sands Advisory Panel 2010; Main 2011) resulted in a program transformation to better manage the monitoring of oil sands influences and the kinds of monitoring undertaken (Environment Canada and Alberta Environment 2012).

RAMP has been shown to demonstrate effects on benthos and water quality, for example, at Tar River and Fort Creek in northwest Canada (Hatfield et al. 2016). Nonetheless, the RAMP approach has been criticized for several reasons. For example, sampling in RAMP was considered to be too infrequent for some environmental components (e.g., water quality) and therefore likely to miss important events (e.g., Kelly et al. 2009, 2010); Further, benthos sampling of riffles with Neil–Hess cylinders has been criticized as inappropriate; several reviews have argued that benthos should be collected with traveling kick protocols (Main 2011). Other criticisms focused on insufficient benthic sampling at some RAMP projects; for example, the absence of benthic sampling locations in the main stem of the Athabasca River (Main 2011). However, RAMP purposely did not survey the mainstem of the Athabasca River because it was considered unlikely to detect the effects of oil sands operations (Hatfield et al. 2009; RAMP 2011). Further, some criticisms of RAMP projects argued that fish population monitoring (i.e., with slimy sculpin) was neither extensive enough (not enough sample locations) nor frequent enough (Main 2011). Other challenges and perceived problems with RAMP are described by Main (2011). Like the EEM programs for metal mining and pulp mills, RAMP reports (e.g., Hatfield et al. 2015) specified values for biophysical responses that were used to judge variations. Unlike the established EEM programs for metal mining and pulp and paper mills, RAMP only involved surveillance monitoring and provided no mechanism (tier) for investigating root causes of any observed variations in the aquatic environment (Miskimmin et al. 2010; Oil Sands Advisory Panel 2010: RAMP 2011).

EEM is an effects-based monitoring program that needs to be a part of any adaptive management process that manages project-related changes to the environment (Arciszewski and Munkittrick 2015). The primary EEM challenges relate to study design, but they also include the setting of expectations for biological response values (i.e., triggers) that justify and prompt a change in monitoring intensity, focus, or questions (i.e., tiers) within EEM.

This paper proposes a template for designing and implementing an EEM program to integrate input from the project-planning stage and predevelopment baseline data collection and triggers that prompt subsequent tiers of monitoring and related activities (e.g., investigation of cause, investigation of extent). The general methodology is anticipated to be applicable to operators of mines, mills, municipal (or other) wastewater systems, and other facilities that influence lake, stream, and river environments. This review discusses the historical development of numeric and narrative triggers as applied to chemical (water and sediment) and biological (plankton, benthos, fish) endpoints and provides an overview of historical study design issues, including what to sample, where to sample, and how frequently to sample. The review demonstrates how conventional control-impact study designs and reference

condition approach (RCA) designs are variants of the same reference–exposure theme and how they can be integrated for a more thorough assessment. In addition, this paper discusses how EEM tiers influence the temporal-spatial statistical design options, the use of triggers, and how the information gained through EEM could be used in an adaptive management context. This paper does not identify preferred chemical, physical, or biological variables for monitoring because those decisions originate in the risk-assessment–based conceptual site model (Dubé and Munkittrick 2001; Ankley et al. 2010) and associated valued ecosystem components, both being project-specific aspects that are handled elsewhere (e.g., Munkittrick and Dixon 1989a, 1989b; Kilgour et al. 2005).

EEM TIERS

The Canadian EEM program was initiated to determine if regulated concentration limits for effluent parameters (under the Fisheries Act) were adequate to protect fish and fish habitats (Ribey et al. 2002; Walker et al. 2002, 2003). As the pulp and paper mills and metal-mining EEM programs matured, flowcharts and decision trees were constructed to describe the major steps (or tiers) and critical decisions in EEM (e.g., Glozier et al. 2002; Ribey et al. 2002; Kilgour et al. 2005, 2007), including data interpretation issues (Lowell et al. 2002; McMaster et al. 2002; Munkittrick et al. 2010) and challenges associated with the investigation of cause (Hewitt et al. 2005; MacLatchy et al. 2010) and the investigation of solutions (Kovacs, Martel, et al. 2007).

Canadian EEM programs include 6 generalized tiers (Figure 1). Tier 1 is baseline period monitoring, which is carried out prior to a project (or intervention) being approved. That is, baseline assessment studies are often initially conducted to describe the existing environment to support a pathways-based risk assessment. However, baseline period monitoring has not been routinely integrated into Canadian EEM programs because many of the EEM programs were implemented to evaluate effluents at existing facilities, making it impossible to collect baseline period data.

Tier 2 is surveillance monitoring, normally the first study undertaken after project development. Surveillance monitoring is used to determine if there are nontrivial (discussed in the section on triggers) changes in biological responses in a nearfield exposure area. Tier 3 is confirmation monitoring, which establishes whether nontrivial effects observed in tier 2 were evident a second time if sampling was repeated. If nontrivial effects are confirmed, the focus on monitoring shifts to tier 4, the investigation of cause (IOC), which is carried out to determine the cause of the nontrivial effects. Once the cause has been identified, focused (extent) monitoring is initiated in tier 5 to quantify the spatial extent of the nontrivial changes, an exercise that helps inform if the changes warrant mitigation. Arciszewski et al. (2017) have

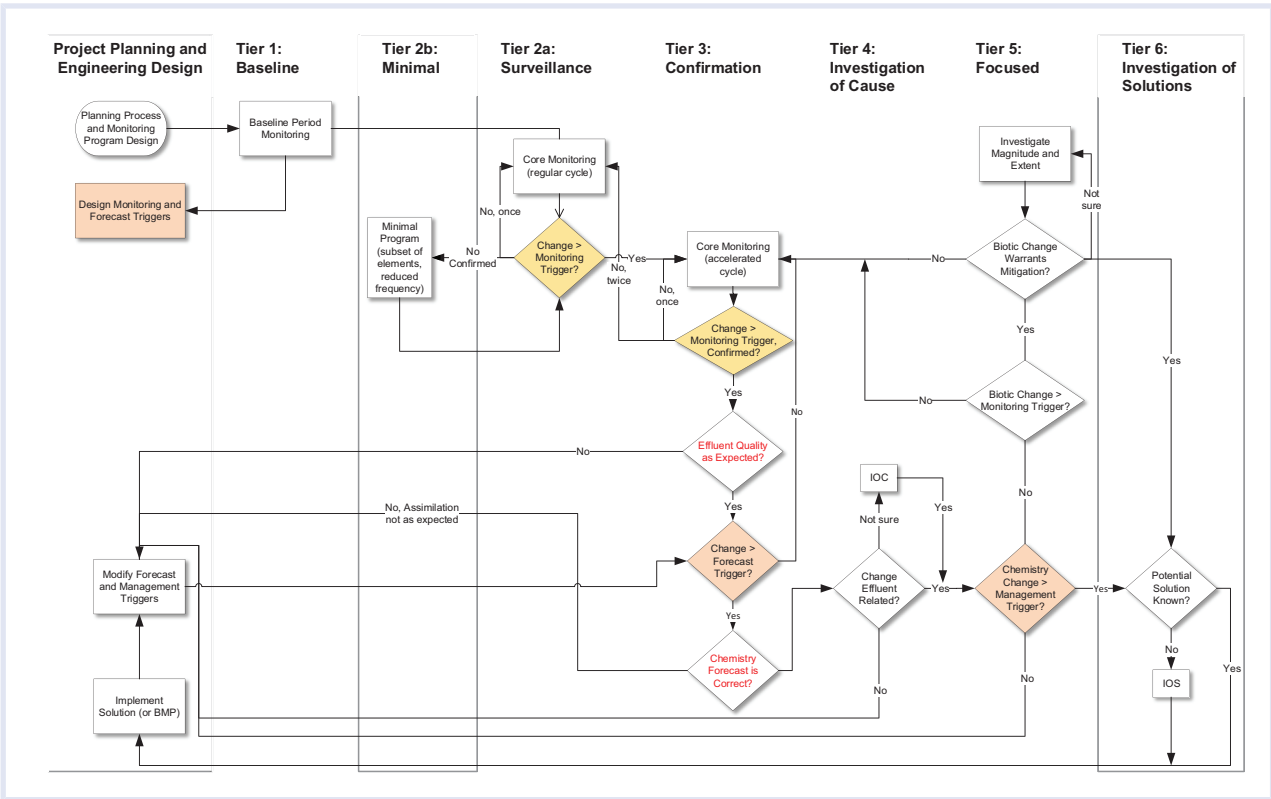

Figure 1. A flowchart illustrating the relationships in a tiered and triggered monitoring system applicable to assessment of effluent release into an aquatic environment. Baseline, forecast, and management triggers are defined as part of the project-planning process. IOS = investigation of solutions; IOC = investigation of cause.

proposed that extent monitoring should be carried out prior to IOC, which reflects a sequence in which the extent determines whether IOC and subsequent mitigation are warranted. However, the IOC confirms that we understand the effects of the project (effluent) on the receiver, and this knowledge provides an opportunity to refine the project's environmental effects predictions in an adaptive management context. Additionally, tier 5's focused (extent) monitoring will provide information that is used to plan and implement mitigation measures. Tier 6 is investigation of solutions, which focuses on an assessment of options that will most likely mitigate the observed environmental effects.

Although often treated as a stand-alone process, EEM is enhanced by connecting it to the initial project-planning and engineering-design phase, which typically involves an EIA or a similar process (Kilgour et al. 2007; Arciszewski et al. 2017). The EIA is the process in which valued components are identified and project environmental effects are predicted based on an analysis of predicted effects pathways; both of these activities are needed to set triggers used to move between tiers in EEM. Additionally, EEM logically connects to an adaptive management (AM) process in which the EEM investigation of cause, extent monitoring, and investigation of solutions tiers support an AM process in which a preferred solution is implemented and evaluated with adaptive monitoring (e.g., see CNSC 2014; Arciszewski and Munkittrick 2015; Arciszewski et al. 2017). The connections between planning and design (EIA), EEM, and AM are logical steps in an EEM for a new project as opposed to traditional EEM programs for an existing facility. When these 3 processes are connected, the planning and design process establishes a number of key monitoring design parameters that are required in the EEM (Beanlands and Duinker 1983; Duinker 1989; Smith 1991). EEM involves the implementation of the monitoring program, and AM is initiated if it is necessary to remedy any observed nontrivial environmental effects.

STUDY DESIGNS IN EEM

The nature of the baseline and surveillance monitoring tiers is determined in part by the physical setting of the facility and associated discharge (e.g., discharge to a stream, lake, or marine embayment; see Glozier et al. [2002] and Environment Canada [2012] for details). By contrast, the study design (also known as the experimental design or sampling design) is often determined by answers to a series of relatively simple questions (e.g., Green 1979; Quinn and Keough 2002; Underwood 1994). Generally, 1 of the first questions is whether any data are available from the area before the facility was operational. These data are called "before" data, and they describe environmental conditions before the facility was constructed (as part of the initial site assessment; e.g., Jensen 1973). If an EIA was conducted prior to the construction of the facility, then suitable "before" data will potentially have been collected (Beanlands and Duinker 1983; Smith 1991). If "before" data are available and we also know when and where the facility and associated discharge are located, then the existence of 1 or more suitable "control" or reference areas will determine the statistical study design that can be applied.

Decision tree

A modified version of Green's (1979) decision tree distinguishes 5 different statistical study designs applicable to the assessment of the release of treated liquid effluents from a point source (Figure 2). Each design describes a different sampling strategy that follows from the availability of "before" data (i.e., baseline period), "control" areas (i.e., baseline or reference areas), and so on. Unfortunately, the designs are not equivalent with respect to inferences of effluent-related cause and effect, with some designs being considerably more robust than others (e.g., Mellina and Hinch 1995; Nemec 1998). At issue is the ability to infer the cause of any significant change revealed by tier 2 surveillance monitoring (i.e., to establish cause and effect; Anderson 1998; Schwarz 1998).

The 5 designs considered here are the following: (1) before vs. after (BA); (2) control vs. impact (CI); (3) before-after-control-impact (BACI); (4) gradient; and (5) RCA. Each is discussed in the sections bellow.

Before-after design

With a before-after (BA) design, data are collected from the area below the effluent release point both before and after effluent release has begun. Samples collected before the effluent was discharged will be compared to samples collected after commencement of the discharge. A nontrivial difference between before and after periods provides evidence that a change has occurred. However, BA designs have several significant challenges. First, changes over time in an exposure area may be natural (Underwood 1994) and unrelated to the intervention (e.g., effluent release). Having multiple years of baseline data, therefore, can be important. Second, projects may have only a couple of years between concept and implementation, i.e., the period when baseline data can be collected before the effluent discharge commences. Further, the physical location of the effluent release point can be expected to change several times during project design, such that baseline data collected even a few years before the effluent is released may not be relevant to assessing the influences of the discharge. Paleolimnological data may provide an alternative means of estimating the normal baseline condition (Bennion et al. 2011); however, the receiver needs to be depositional in nature (i.e., lake, large bay of a river, marine offshore) with minimal disturbance of surficial sediments, the chemicals cannot be those that migrate, and the fossil record is complete only for those taxa with calcareous (clams and snails, Kerr-Lawson et al. 1992), siliceous (diatoms; Smol and Stoermer 2010), or chitinous (i.e., chironomids; Walker 1987) exteriors. As a result, BA designs have challenges, and the results from these kinds of studies need to be interpreted cautiously. Like other designs, BA designs can be augmented with regional reference data to put the observed changes in a broader context (Kilgour et al. 2017).

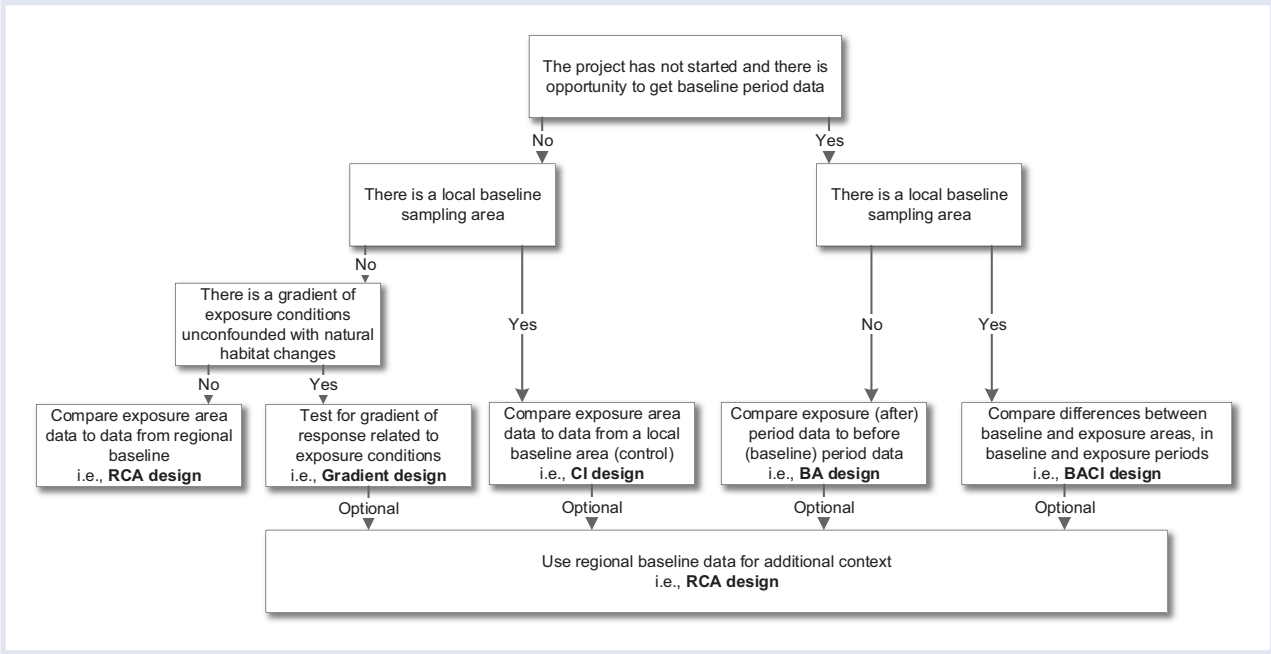

Figure 2. The sampling-design decision tree (modified from Green 1979) for choosing among 5 experimental designs that could be used in the surveillance-monitoring step in EEM.

Control-impact design

The control-impact (CI) design compares observations from the exposure area to 1 or more reference areas during the exposure period. This design assumes that, in the absence of a discharge, responses in reference and exposure areas are the same and that differences between reference and exposure areas indicate that a change in the exposure area has occurred. The design is compromised, however, because differences in responses between any 2 sampling areas are virtually guaranteed because of physical, chemical, or just natural variability (Underwood 1991, 1994). Despite this limitation, the CI design has been extensively used in Canadian EEM programs to evaluate existing discharges (i.e., pulp mill and metal mine effluents). Because 2 locations naturally differ in chemical and biological measures, practitioners often collect data from multiple reference areas in anticipation that data from multiple reference areas more fully characterize the variation in reference conditions (Environment Canada 2012) and reduce the likelihood of detecting differences that erroneously indicate that changes in response in the exposure area have occurred. These CI designs with multiple reference areas, however, still suffer from the possibility that responses in the exposure area have always differed from responses in all of the reference areas. Again, and like other designs, CI designs can be augmented with regional reference data to put the observed changes in a broader context (Kilgour et al. 2017).

BACI design

The BACI design overcomes some of the challenges of the BA and CI designs (Green 1979; Mellina and Hinch 1995). In the BACI design, changes in the differences between reference and exposure areas between before and after periods are used to demonstrate the effects of the perturbation (e.g., effluent release). Despite being described as the "optimal" design, the BACI design also has problems. Underwood (1991, 1992, 1994) and others (e.g., Stewart-Oaten et al. 1986, 1992) have noted that multiple reference areas are needed to accommodate the fact that reference areas naturally differ and reference areas also differ in their inherent variability in space and time. Thus, the use of multiple reference areas provides estimates of variation among reference areas that can be incorporated into comparisons to ensure that observed changes between reference and exposure areas are evaluated relative to natural variability among reference areas. That is, multiple reference areas provide a more realistic estimate of normal background conditions and greater confidence in the assessment of the exposure-area data.

Gradient design

A gradient of effluent exposure conditions, unrelated to natural habitat features (e.g., channel slope), provides a different opportunity to test for effluent-related effects (Environment Canada 2012). In a gradient design, sampling areas are situated at increasing distance from the point of discharge. The increasing distance and associated dilution cause a gradient of exposure. A correlation between response (biological), exposure (chemical concentrations), distance from source provides the evidence of an effluent-related effect. A challenge with gradient designs is that many habitat factors often vary over distance, such as water depth, substrate texture, and flow volume (in riverine systems). As a result, correlations between the biological response and distance from the source can occur naturally.

This design can be used in the absence of temporal or spatial baseline data, but the interpretation will benefit if those data are available, particularly if there are changes in natural habitat features with distance from source (e.g., Kilgour et al. 2007). Like the other designs, results from gradient designs can be put into a broader perspective if the results are compared to regional baseline data (i.e., RCA data).

Reference condition approach

In the Canadian metal-mining, pulp, and paper EEM programs, "before" data were often unavailable, and this lack of information limited the types of study designs that could be used. As a result, Environment Canada acknowledged that the RCA (Figure 2) could be used as an alternate design in EEM (Glozier et al. 2002; Environment Canada 2012). The RCA typically compares data from a single exposure area with data from a collection of minimally impacted reference (or control) areas (e.g., Bailey et al. 2004; Bowman and Somers 2005). Variability among reference areas is used when comparing the exposure area with the average of the reference areas, so this design differs statistically from the other study designs. The RCA has gained popularity because it provides a solution to the problems of missing "before" data and/or the absence of an appropriate upstream "control" area. The need for multiple reference areas increases the cost of the RCA design, but proponents can share reference-site data if the sites are reasonable matches for exposure areas. RCA designs and the analysis of their data result in "modeled" estimates of what the response in the exposure area should be (Bailey et al. 2004). The RCA designs, however, normally use data from reference locations that are geographically removed from the exposure area and thus provide a regional reference expectation (Kilgour et al. 2017). Combining RCA model predictions with the results from a BACI design can be very informative: the BACI design can illustrate the magnitude of change attributable to the project, whereas the RCA data provide an indication of regional normal ranges that can be used to evaluate local changes in a broader context (Figure 2; Kilgour et al. 2017).

TRIGGERS

Rationale

Most EEM programs are designed to test for "a significant change" in chemical, physical, or biological responses, where "a significant change" is regarded as a statistically significant nonzero difference from an expected reference condition. Canadian metal mines, for example, are required to statistically test for significant differences between means of observations from reference and exposure areas under classic CI designs (Environment Canada 2012). When relatively simple study designs are used (like BA and CI designs; e.g., Green 1979; Duinker 1989; Smith 1991), nonproject-related influences (i.e., unrelated to exposure to effluent or the project) can cause changes from before to after or differences between reference and exposure, and they therefore complicate the interpretation of observed differences and changes. Detection of statistically significant change can be guaranteed if enough samples are collected (Underwood 1991; Mapstone 1995). In the metal-mining EEM program, differences between reference and exposure areas are very likely given the large number of response variables that are considered in the assessments (i.e., 9 different fish population responses are evaluated on each of 2 sexes on each of 2 species, Huebert et al. 2011). Further, given that EEM programs are designed to assess the effects of an intervention (i.e., effluent release) on an ecological response, the intervention will de facto cause change in the physical and chemical environment that will de facto result in change in biological responses no matter how trivial those changes are (Underwood 1994). As a result, the inability of an EEM program to detect a change is simply a problem of sample size (e.g., Osenberg et al. 1994; Underwood 1994; Mapstone 1995) and is not related to whether or not a change occurred: a change did occur, regardless of how small and statistically undetectable the change was. The logic of EEM programs is, therefore, improved when programs are designed to test for changes of a prespecified magnitude, rather than testing for "any" change, including those that are trivially small (see also McBride et al. 1993, Hanson 2011). Herein, we use the term "trigger" to refer to a value of a response variable that can be used to justify a continuation of the current monitoring tier or a change in monitoring tier (Arciszewski et al. 2017).

Types of triggers

There are at least 5 different types of triggers that can be used to precipitate changes in the management of a monitoring program: (1) performance; (2) compliance; (3) monitoring; (4) forecast; and (5) management triggers (see Figure 3). Performance and compliance triggers operate at the facility scale and relate to critical concentrations of constituents associated with emissions (effluent) from the facility. Performance triggers are levels set by process engineers that indicate whether the "process" is working as designed. Exceedance of performance triggers (also

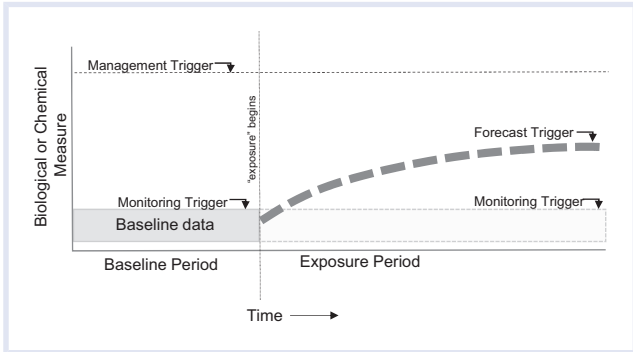

Figure 3. Conceptual illustration of the relationships between monitoring, forecast, and management triggers, as they might be applied to a biological or chemical indicator in an environmental effects monitoring program.

termed action levels by some companies) results in a review of the engineering process, potentially with adjustments, to ensure that the facility process is operating per specifications. Continued exceedance of performance triggers may result in an update of the effluent-model forecast and reestimation of the potential for effects in the receiver if engineering process solutions are deemed too difficult, too costly, or otherwise unwarranted. Discharged effluents are nearly always licensed, with the licensees dictating limits that are not to be exceeded without a legal consequence. License limits are here considered compliance triggers. Exceedance of license limits (like those in the MMER for pH, copper, and nickel) normally results in an increased frequency of monitoring to confirm the exceedance, engineering studies to determine the cause, and process modification to eliminate the exceedance. Repeated exceedance of license limits can lead to charges or loss of license, without adequate demonstration of due diligence to address the exceedances. Performance and compliance triggers are not considered further because they are not used in the receiving environment; however, forecast, monitoring, and management triggers apply to the receiving environment.

Forecast trigger. On the basis of facility operation knowledge, anticipated emissions (effluent) quality, and dilution ratios in the receiving environment, engineers normally provide a prediction (or forecast) of anticipated chemical concentrations in the receiving environment. The predicted environmental concentrations can be used as a forecast trigger in order to evaluate observed chemical concentrations in the receiver. Through the EIA (or a similar) process, forecasts of anticipated changes to ecological receptors in the environment may be modeled, although those predictions are normally less quantitative and more narrative than predictions derived for water quality variables. Because forecasts for biotic variables are often vague and general, monitoring triggers are frequently used as surrogates for forecast triggers (but without being stated as such). AREVA Canada, for example, forecast that uranium mine effluent would cause "negligible" effects on benthos in Collins Creek in the downstream exposure area (AREVA 2009). Through multiple cycles of EEM reporting, AREVA has adopted normal ranges of reference conditions to define monitoring and forecast triggers (though those terms are not used in their technical documentation; AREVA 2009, 2012). Forecast triggers should include consideration of when, where, and potentially for how long the changes will occur. Expressed in those terms, forecast triggers define a priori when we expect to see change and where. Regardless, observed responses in the exposure area that exceed the forecast triggers indicate that (1) forecast conditions have changed; or, (2) that our understanding (i.e., the model) of the process and receiver were inadequate. Both instances require that the system (engineering process and receiver) be studied further, the forecast model recalibrated, and the recalibrated model used to reforecast future conditions (Figure 3).

Monitoring trigger. Baseline data for water, sediment, and biota collected during implementation of BA, CI, BACI, gradient, or RCA designs provide the data to derive monitoring (baseline) triggers. Monitoring triggers are typically based on some measure of background variability or the normal range of variation of the monitoring data (Arciszewski et al. 2017). The concept of the normal range has been adopted as a generic approach for deriving site-specific triggers for surveys of water quality (Chambers et al. 2012), sediment quality (Thompson et al. 2005; Kilgour et al. 2017), and benthic macroinvertebrates (Lowell et al. 2002; Bailey et al. 2004, Parsons et al. 2010; Environment Canada 2012). Data used to derive monitoring triggers based on the normal range typically originate from site-specific or regional reference areas lacking perturbation (Wissmar 1993; Roux et al. 1999; Hanson 2011) or lacking influence from the intervention being assessed (i.e., effluent). The normal range can be estimated from a collection of random samples from a number of minimally influenced reference areas (Wissmar 1993; Roux et al. 1999; Hanson 2011). Statistically, measurements on a collection of observations are typically described by an average (e.g., the mean of the reference data, or $\bar{x}_r$) and the variation around that mean (e.g., the standard deviation or SD). In situations where the distribution of the observations around the mean is reasonably described by a normal curve, then approximately 95% of the observations should fall within the region under the curve enclosed by $\bar{x}_r \pm 2\text{SD}$, which then defines the normal range for any variable (Kilgour et al. 1998). Because these simple statistics can be calculated for any variable regardless of the measurement units, the normal range is a generic statistic that can be used for most physical, chemical, and biological variables. The exceedance of a monitoring trigger implies that a nontrivial change has occurred and that change should be investigated as a potential early warning of unexpected changes in the monitoring results.

Management trigger. Management triggers here are desired values for water quality, sediment quality, physical conditions, and biota as determined through an EIA or a similar process (e.g., regional planning; Government of Alberta 2012) by stakeholders such as industry, government agencies, and the public (IAEA 2010; Dubé et al. 2013). Exceedance of management triggers indicates that the overall objectives were not being met, with different potential outcomes depending on whether the response measure was chemical, physical, or biological. Chemical and physical variables generally indicate the quality of habitat for biota. Exceedance of a chemical guideline, for example, does not guarantee that triggers for biota will be exceeded. Exceedance of triggers for biota are more critical and may indicate that overall objectives for resource protection have not been met. Like forecast triggers, management triggers normally include a consideration of the location where the trigger applies. As such, management triggers define a priori the geographic point where change would be deemed unacceptable (see 2WE Consultants Ltd. [2002] for an

example of setting a spatial boundary on management triggers). Management triggers may include water-use and protection-based limits derived from water-management goals at specific locations, ranging from nondegradation (no measurable change from background) to industrial water use (e.g., CCME 2003). Management triggers may also be defined as a change far beyond a monitoring trigger that may not necessarily require confirmation (Arciszewski et al. 2017).

Water and sediment quality guidelines are often used as management triggers. For example, Policy 1 surface waters in Ontario are watercourses where the concentrations of chemicals are currently below provincial water quality objectives (OMOEE 1994). Water quality objectives in those Policy 1 situations are used as management triggers in planning exercises such as subwatershed studies (OMOEE 1993). Policy 2 surface waters are watercourses where chemical concentrations exceed water quality guidelines, and the existing condition becomes the objective (or the management trigger; OMOEE 1994). Setting management triggers to be greater than water quality objectives or guidelines is not unique to Ontario. British Columbia (BC) uses both water quality guidelines and science-based guidelines for setting environmental "benchmarks" (triggers) under the BC Environmental Management Act for permitting mines (BCMOE 2016). Alberta Environment and Parks (2015a, 2015b) has set water quantity and quality targets for the Lower Athabasca River, based on analyses of historical and baseline data and in consultation with government, industry, and nongovernment organization stakeholders. Those targets are equivalent to the management triggers described here. In the case of the Lower Athabasca River, management triggers for water quality are based on normal ranges of historical data and in some cases exceed water quality guidelines (e.g., P) and in other cases are lower than water quality guidelines (e.g., chloride).

Many water and sediment quality guidelines (e.g., CCME 2003; USEPA 1986; Gaudet et al. 1995) were historically based on the assimilative capacity of the receiving environment and can be used to interpret observed water and sediment chemistry data. Water quality guidelines are now based on the results of toxicity tests in controlled laboratory environments and often use species sensitivity distributions to compute a lower concentration that can be anticipated to protect some fraction (normally 95%) of aquatic biota (e.g., Cormier and Suter 2013; Zajdlik 2016). The CCME (2007), for example, requires that there be laboratory toxicity data for at least 7 aquatic species (2 fish, 1 amphibian, 1 zooplankton, 1 benthic invertebrate, 1 algae, and 1 more) and that the data must meet minimum standards for robust methodology before they will be included in the derivation of a guideline.

Sediment quality guidelines are more variously determined (Landrum 1995). In Ontario, sediment quality guidelines were based on a screening-level concentration methodology in which low-effect levels are defined as the concentration that can be tolerated by 95% of benthic taxa and severe effect levels can be tolerated by 5% of benthic taxa (Neff et al. 1986; OMOEE 1993). Concentrations of water or sediment quality parameters that exceed toxicity-based guidelines pose potential risks to aquatic biota, whereas concentrations that do not exceed guidelines pose less obvious potential risks to biota. Water and sediment quality guidelines may be relevant to effect-based responses, depending on the measured analyte (the stressor) and the measured response (or not, it depends; see Gaudet et al. [1995] and Roux et al. [1999]). Site-specific water quality values can be derived (CCME 2003; den Besten et al. 2003), and they may more accurately quantify likelihood of harm to the biological response, again depending on the analyte and response. Species-specific toxicity reference values can be used in place of more generic quality guidelines, if the interest is in computing a risk quotient specific to the measured effect variables (e.g., fish, benthos).

Statistical comparisons using triggers

Comparison of the response data to a trigger value will inevitably require formal statistical procedures, with the specific method depending on how the trigger was derived. Management triggers are typically "fixed" values such as water or sediment quality guidelines (some guidelines vary depending on a covariable such as water pH or hardness, but the guideline is essentially a known value that is applied to a site). Comparison of data from an exposure area relative to fixed (known) values can involve a 1-sample analysis of variance or 1-sample *t* test that determines the likelihood that the "mean" concentration in the exposure area was lower or higher than the trigger value. There may also be an interest in testing whether a certain fraction (say 5%) of observations from the exposure area exceed a water or sediment quality guideline. This assessment can be addressed with tolerance limits for an upper percentile (e.g., Smith et al. 2003; Kilgour et al. 2017) applied to the exposure-area data or through bootstrapping to derive confidence limits for an upper boundary such as the 95th percentile (Anderson and Thompson 2004).

Other triggers may contain uncertainty in their estimated value. Forecast triggers, for example, are based on numeric models that reflect our understanding of environmental processes. Forecast models contain uncertainty (Borsuk et al. 2002), which should be reflected in the future forecast values (see Figure 3, illustrating that the forecast is a range of values for any given period in the future). Comparison of the observed data to the forecast values should account for the variability of both the forecast and observed data and consider what type of exceedance is important (e.g., by the mean or some specified proportion of the data). Again, tolerance limits and bootstrap procedures provide additional options for the comparison.

Monitoring triggers are typically derived from 1 or more sets of reference-area data, with sometimes considerable uncertainty around the trigger value. In the Canadian EEM experience, the critical effect size for benthos is the estimated normal range for samples of reference-area data (i.e., $\bar{x}_r$), where both the mean and SD are estimated with uncertainty. Likewise, the critical effect sizes for fish population

parameters (e.g., a 10% change in condition factor or 25% change in liver and gonad size) are expressed relative to reference-area data. The critical effect sizes for benthos and fish, therefore, are not fixed values, because they depend on samples of reference data and there will be residual uncertainty in the precise value given sampling error, etc. For triggers based on reference-area data, special analytical approaches are required to demonstrate that the observed change was equal to or greater than the critical effect size (McBride et al. 1993; Hanson 2011), potentially involving noncentral analysis of variance (Kilgour et al. 1998, 2017; Bowman and Somers 2006) or similar nonparametric procedures (Murphy 1948).

INTEGRATION OF TRIGGERS WITH TIERS

Project planning and engineering design

A project-planning and engineering-design process (such as an EIA or similar exercise; see Spellerberg 2005) provides the project description, forecasts of anticipated environmental (chemical, physical, biological) changes in the environment, a risk assessment to estimate effects to receptors, and a cumulative effects assessment to integrate the effects of the project with the effects of other historical, current, and potentially future disturbance activities (Dubé et al. 2013). The first step in an EIA processes is the creation of the project description (Beanlands and Duinker 1983; Smith 1991; Spellerberg 2005), which outlines the nature and benefits of the project, its feasibility, operational life cycle, and closure plan. This information is typically supported by a scoping or screening study that establishes the spatial extent of the project and collects baseline information on the physical, chemical, and biological components that may be altered by the project. Public consultation, part of the environmental assessment process, identifies valued ecosystem components (VECs) for which there is a desire to manage the environmental conditions and therefore monitor the project and environmental performance (Duinker and Beanlands 1986). EEM practitioners in consultation with government agencies identify the relationships between EEM endpoints and VECs.

Information from the EIA documentation and scoping steps is used to model and predict anticipated environmental alterations associated with the project, based on predicted effects pathways as well as magnitude and significance assessments (Spellerberg 2005). Where these predictions suggest alterations that may impact the VECs, a series of mitigation measures is proposed to ensure that potential impacts are minimized. The prediction and mitigation step is followed by a project management and monitoring plan and an auditing plan. The project management and monitoring plan outlines the activities associated with routine operation of the facility, including monitoring activities to ensure that routine operations are functioning as expected (per performance triggers). Once compiled, the EIA documentation is submitted to the appropriate governance body in order to receive authorization to proceed with the project. Following authorization, the project moves to the formal permitting stage where triggers and limits are set, for example, as part of the effluent discharge permit. The EIA (or similar process), therefore, is the time when triggers are set in order to judge project performance and compliance. The EIA (or similar) process is also the time when forecast, monitoring, and management triggers can and should be established.

The Terra Nova oil and gas drilling project off the shore of Newfoundland (Deblois et al. 2014) provides an example of the EIA stage providing testable predictions for a follow up EEM program. In this situation, a literature review was used to support the forecast of the potential effects of the development on water quality, sediment quality, and benthos. Historical data for similar projects indicated that offshore drilling would likely result in drill cuttings (suspended sediments and associated chemicals) depositing as far as 15 km from the source, with the greatest deposition occurring within 5 to 10 km. On the basis of the literature review, effects of those drill cuttings on benthic invertebrates were expected to be "mild a few hundred metres away from drill centres, but fairly large in the immediate vicinity of drill centres" (Deblois et al. 2014). The forecasts for physical and chemical alteration, therefore, were reasonably quantitative, allowing for direct numeric comparison (Deblois et al. 2014). The forecast for benthic invertebrate community composition was much less quantitative and resulted in an assessment using normal ranges for reference data (i.e., baseline period data, in addition to data from distant reference areas) in order to judge where effects on benthos had occurred (Paine et al. 2014).

Tier 1: Baseline monitoring

Because baseline monitoring triggers are routinely based on reference data, baseline period reference data will ideally be collected during the development of the EIA in both reference (unexposed to effluent during operaion) and exposure (exposed to effluent during operation) areas. Monitoring during the EIA stage is rarely designed with the recognition that the collected data may, could, or should be used in a BACI context to assess environmental performance. It is precisely during the EIA stage, however, that the opportunity exists to collect the required data (Kilgour et al. 2007).

For new facilities, baseline information collected during the EIA can provide critical information with respect to defining the spatial and temporal extent of the discharge; the physical, chemical and biological nature of the area before the discharge was initiated; the existence of 1 or more minimally influenced reference or control areas not affected by the discharge; and the existence of potential confounding factors (including existing changes to baseline conditions; e.g., upstream discharges) that may affect data interpretation (Beanlands and Duinker 1983). Additionally, the baseline EIA data may provide estimates of the normal range of variation for the physical, chemical, and biological components of the aquatic ecosystem before the facility was constructed. This information can be used to develop site-specific triggers for

chemical and biological measures, where exceedances can be used to prompt movements through the tiers in the EEM flowchart.

The collection of appropriate baseline period data during the EIA stage is not routine but has some good examples. Per the Terra Nova example off the shore of Newfoundland, Deblois et al. (2014) collected baseline data for 2 years in areas that were subsequently used to support surveillance monitoring. Monitoring under RAMP in Alberta's oil sands region was designed such that baseline data were supposed to be collected for 3 years (for a tributary) prior to oil sands extraction activities adjacent to that tributary (Hatfield et al. 2009).

Tier 2: Surveillance monitoring

The first step in EEM is the design and implementation of a surveillance-monitoring program (Figure 1). The data for surveillance monitoring will often be collected per the sampling design that was developed and implemented during the baseline period and with a core set of components. Although standard protocols have been developed for EEM (e.g., Environment Canada 2012), each facility is often located in a unique physical setting, so methods will tend to vary from location to location, as will the physical layout of sampling stations, areas and zones, etc. Consequently, appropriate field and laboratory protocols must be selected to generate unbiased samples that accurately represent environmental conditions at each facility (Glozier et al. 2002). Surveillance-monitoring results are compared to baseline monitoring triggers (Figure 2). Exceedance of baseline triggers justifies a transition to tier 3 confirmation. Because it is difficult to quantitatively predict outcomes for biological variables, baseline monitoring and forecast triggers for biotic variables may be the same value.

Tier 2 surveillance monitoring is carried out to confirm that the predictions from the project-planning exercise were correct. As a result, monitoring should be continued at some frequency until the burden of evidence suggests that the forecast was correct or incorrect. Surveillance-monitoring programs in Canada span a range of frequencies. For example, biological monitoring in the oil sands region occurs annually in tributaries to the Athabasca River where there is a heightened concern of potential and future change (e.g., Steepbank River, Muskeg River, Mackay River, Tar River; Hatfield et al. 2009), whereas monitoring occurs at a reduced frequency for tributaries where there is less concern of potential change (e.g., Clearwater River; Hatfield et al. 2009). The frequency of surveillance monitoring under the Canadian EEM program has been every 3 years, with the potential for a reduction in monitoring frequency (to every 6 years) if a lack of significant change is confirmed. The rationale for surveillance monitoring (Figure 1) frequency can be based on concern level (per the RAMP; Hatfield et al. 2009) or on anticipated ecological response times. Freshwater benthos have life cycles that vary from a few months to a few years. Monitoring benthos every 2 to 3 years, therefore, has some logic. By contrast, the time to maturity for freshwater fish varies from 2 to 10 years (Scott and Crossman 1973), depending on the species and local conditions. As with benthos, monitoring at a frequency of every 2 to 3 years has a sense of logic. A reduction in monitoring frequency is triggered in Canadian metal-mining EEM programs if trivial or nonsignificant differences are observed in consecutive surveys. A reduction in monitoring components may also be justified.

Tier 3: Confirmation monitoring

Regardless of whether observed changes exceed monitoring or forecast triggers, the observed changes should be confirmed (i.e., confirmation monitoring; Glozier et al. 2002; Lowell et al. 2002) before any changes to operations are justified. In this tier, components used in surveillance monitoring are used with a goal to establish whether the observed initial changes are evident a second (or third; Arciszewski and Munkittrick 2015) time. Various quality assurance rules can be applied to data collected repeatedly over time, and process control decisions can depend on the frequency and magnitude of exceedances (Westgard et al. 1981). If confirmation monitoring fails to confirm change, the EEM program logically returns to surveillance monitoring, and the EEM program sampling may be reduced in frequency or delayed until there are demonstrated changes in effluent quality (Figure 1). For example, the Canadian metal-mining EEM permits a reduction in surveillance monitoring frequency from every 3 years to every 6 years, if there are no large changes in 2 successive cycles (Environment Canada 2012). If confirmation studies fail to confirm effects but document different effects, then confirmation studies are conducted on the new effects. Changes that exceed monitoring triggers need confirmation to ensure that the observed change was not caused by unusual factors (Lowell et al. 2002). Changes exceeding forecast and management triggers may be severe enough that confirmation needs to be done immediately or without delay. Confirmation of observed changes, however, triggers IOC to determine if the changes are effluent related (Figure 1).

Tier 4: Investigation of cause

Tier 4 IOC is initiated by confirmed exceedance of the forecast triggers and when the cause of the exceedance is uncertain. Any unexpected biological or chemical change that exceeds the forecast triggers indicates that our understanding of the system is incomplete (Dubé et al. 2013). Data from the IOC will contribute to our understanding, will inform present and future projects (through their EIAs), or will assist in adaptively managing conditions in the future (Arciszewski and Munkittrick 2015). IOC can be informed by patterns in the results from surveillance, confirmation, and focused monitoring. For example, patterns in the data might help to reveal dose–response relationships between the effluent concentration and biological responses (Hewitt et al. 2005; Davies and Jackson 2006). Additionally, different patterns among water quality and fish population responses (Munkittrick and Dixon 1989a, 1998b; Gibbons and Munkittrick 1994) and for benthic communities (Pearson

and Rosenberg 1978; Environment Canada 2003) can be used to infer the causes of the observed change. In EEM, 3 types of response patterns have been commonly observed: (1) increased fish condition factor, liver size, and gonad size with increased abundance and richness in the benthos; (2) increased condition factor, increased liver size, and decreased gonad size in fish, but no consistent benthos response; and (3) decreased condition factor, decreased liver size, and decreased gonad size in fish with decreased abundance and richness in the benthos. These 3 patterns were attributed to (1) the effects of nutrient enrichment or eutrophication, (2) metabolic disruption, and (3) the toxic impacts of 1 or more contaminants (Hewitt et al. 2005). Depending on the response pattern, the IOC tier can benefit from controlled laboratory exposures to diluted effluent, detailed chemical analyses, and experimentation to determine the mode of action and dose–response relationships (e.g., Walker et al. 2005; MacLatchy et al. 2010; Kovacs et al. 2011).

Tier 5: Focused (extent) monitoring

Tier 5 focused (extent) monitoring is prompted by (1) confirmed changes that exceed the forecast triggers in tier 3; (2) confirmed changes that are demonstrated with confidence to be effluent related; and (3) uncertainty that the changes warrant mitigation (i.e., justifying initiation of tier 6 investigation of solutions). A study to explore the spatial extent of change (tier 5) should provide the necessary information to determine if the change warrants mitigation. In tier 5 focused (extent) monitoring, the sampling protocols used in tier 2 surveillance and tier 3 confirmation are used to determine the spatial extent of changes. In many respects, this tier is a mapping exercise in which additional effort is invested in mapping the spatial extent of the observed changes. Focused (extent) monitoring helps to determine whether the forecast was correct and may be done simultaneously with an IOC to inform us of the urgency with which to carry out the IOC and develop a mitigation plan (i.e., tier 6). Available information associated with the modeled effluent plume trajectory and dilution rates will help to situate and select sampling areas in order to map changes relative to forecast triggers and better understand the potential sources contributing to these conditions. Additionally, the resultant fine-scale spatial data will provide opportunities to model interrelationships between the biological, chemical, and physical conditions in an effort to reveal dose–response relationships that will assist in understanding the potential cause(s) of the observed differences.

Often the study design associated with focused (extent) monitoring consists of a gradient design (see Figure 2) with samples collected along a gradient of decreasing effluent concentration with increased distance from the effluent discharge (Glozier et al. 2002; Ribey et al. 2002). If biological endpoints are consistent with the gradient (e.g., Davies and Jackson 2006), these results support the hypothesis that the effluent source is potentially causing the differences observed in the previous tiers. These spatial data, further, can be used to establish an exposure-area response pattern across the suite of fish and benthos metrics in order to identify the potential cause of the observed differences (e.g., Hewitt et al. 2005; Kilgour et al. 2007; Munkittrick et al. 2010). For example, a contaminant issue would be implicated if the benthos abundance and richness metrics were significantly reduced in the exposure area. For the fish metrics, lower condition factor, smaller liver size, and smaller gonad size would further suggest a contaminant issue (Hewitt et al. 2005).

Surveillance monitoring may be designed a priori (i.e., as part of baseline monitoring) to provide the data that will inform the extent of change. The Terra Nova EEM (Deblois et al. 2014), for example, sampled very near the drill centers and along a gradient of exposure to a distance of 20 km from the drill centers to (1) demonstrate effects exceeding normal ranges near drill centers, and (2) illustrate how far from drill centers the effects exceeded the normal-range monitoring trigger. Because the program was carried out every second year for a period of 14 years, the program confirmed effects and through the gradient design, determined extent.

Tier 6: Investigation of solutions

The investigation of solutions (IOS) tier is initiated when management triggers are (or can be anticipated to be) exceeded and when the changes warrant mitigation. Evidence for a future exceedance could be based on a temporal trajectory in a response that is more than the forecast. Tier 6 is often viewed as the first step in an AM cycle (Figure 1). The primary objective of the investigation of solution tier is to identify cost-effective mitigation and remedial strategies to address the causal factors contributing to the observed environmental effects (Kovacs, Hewitt, et al. 2007; Martel et al. 2010). The 3 exposure-area response patterns implicated nutrient addition, metabolic disruption, and contaminants as the typical causal factors (Hewitt et al. 2005). In cases where nutrient enrichment was the primary problem, the solution focused on reducing nutrient concentrations in the effluent. Similarly, the contaminant-response pattern invoked efforts to reduce the effluent concentration of the key contaminant(s). By contrast, the cause of the metabolic-disruption response pattern was often more difficult to isolate and identify (Hewitt et al. 2008; Martel et al. 2010), and work continues to develop methods to identify the causes of metabolic disruption (Hewitt 2011).

In many respects, this tier 6 IOS is principally an engineering exercise and may not involve monitoring of the receiving environment. Assuming that there are in-house adjustments to routine operations and associated effluent, the forecast of conditions for the receiving environment will be updated to reflect the new information gleaned from this tier 6 IOS. Once changes are implemented, the EEM process should logically return to tier 2 surveillance monitoring with the caveat that the expected environmental (chemical and biological) conditions will include the influences of both historical and current effluent quality: the forecast of future

conditions will, therefore, require recalibration and estimation. Teasing out the influences of historical conditions on present-day observations is problematic (Burton et al. 2014) but nonetheless an important element for determining if the current effluent remains a cause of unacceptable environmental conditions.

It may be that complete mitigation is not feasible and that the management triggers are in conflict with operation of the project. As a result, management triggers may need to be reset to a new level in order to allow the project to continue. Tier 6 IOS, therefore, can also consider input and approval from stakeholders in order to reset management triggers. Controlled experimentation may reveal that management triggers for chemical measures were originally set too low and can be set higher (for example, via procedures for setting water quality objectives; CCME 2007) while maintaining overall objectives such as a diverse and functioning aquatic system (e.g., Kilgour et al. 2005).

Adaptive management: Adaptive environmental effects monitoring

Although the post-IOS return to the tier 2 surveillance monitoring may be viewed as a continuation of the EEM program, this step may be seen as a transition to an AM cycle. The basic steps in AM involve (per Holling 1978) (1) assessment of the problem; (2) design of a solution; (3) implementation of the solution; (4) observation of the results; (5) evaluation of the outcome; (6) adjustment of the solution; and (3) implementation of the solution and continued observation to ensure that the problem has been resolved. The EEM tier 6 IOS is similar to the initial AM step, and the investigation of solution is equivalent to designing a solution. In the EEM—Adaptive Management flowchart (Figure 1), once a solution has been proposed and implemented, the adaptive monitoring step follows (see Arciszewski and Munkittrick 2015). The study design and choice of triggers for adaptive monitoring can draw from experiences with the EEM surveillance-monitoring step. If the implemented solution has been effective at eliminating the cause of the observed differences in the original EEM, the initial surveillance monitoring should reveal no significant changes in excess of the forecast, and surveillance monitoring would continue on a regular schedule. However, if the solution failed to adequately mitigate change, then the tier 2 surveillance-monitoring data would exceed the monitoring or forecast trigger prompting tier 3 confirmation and the basic EEM/AM steps would follow in order to identify the cause, associated solution, further mitigation, and ultimately, remediation and restoration.

Acknowledgment—The development of this manuscript was made possible by funding provided by Canada's Oil Sands Innovation Alliance to Kilgour & Associates Ltd. The authors declare no conflicts of interest. The manuscript was reviewed and improved by several anonymous reviewers. We thank them for their comments and suggestions to improve the manuscript.

Data Accessibility—No external data were used or analyzed in this study.

REFERENCES

2WE Consultants Ltd. 2000. A Comparative review of the CRD's environmental monitoring programs for the Clover Point and Macaulay Point Wastewater Outfalls. Victoria (BC): 2WE Consultants Ltd. 39 p.

Alberta Environment and Parks. 2015a. Surface water quantity management framework for the Lower Athabasca River. Edmonton (AB): Alberta Environment and Parks. 88 p. http://aep.alberta.ca/lands-forests/cumulative-effects/regional-planning/documents/LARP-SurfaceWaterQuantity-Feb2015.pdf

Alberta Environment and Parks. 2015b. Surface water quality management framework for the Lower Athabasca River. Edmonton (AB): Alberta Environment and Parks. 52 p. http://aep.alberta.ca/land/programs-and-services/land-and-resource-planning/regional-planning/lower-athabasca/documents/LARP-SurfaceWaterFramework-Aug2012.pdf

Anderson JL. 1998. Errors of inference. In: Sit V, Taylor B, editors. Statistical methods for adaptive management studies. No. 42. Victoria (BC): Ministry of Forests and Range Research Branch. p 68–87.

Anderson MJ, Thompson AA. 2004. Multivariate control charts for ecological and environmental monitoring. *Ecol Appl* 14:1921–1935.

Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD, Mount DR, Nichols JW, Russom CL, Schmieder PK, et al. 2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ Toxicol Chem* 29:730–741.

Arciszewski TJ, Munkittrick KR. 2015. Development of an adaptive monitoring framework for long-term programs: An example using indicators of fish health. *Integr Environ Assess Manag* 11:701–718.

Arciszewski TJ, Munkittrick KR, Scrimgeour GJ, Dubé MG, Wrona FJ, Hazewinkel RR. 2017. Using adaptive processes and adverse outcome pathways to develop meaningful, robust, and actionable environmental monitoring programs. *Integr Environ Assess Manag* 13(5):877–891.

[AREVA] AREVA Resources Canada Inc. 2009. McClean lake operation caribou project environmental impact statement. Saskatoon (SK): AREVA Resources Canada Inc. 20 p. http://environment.gov.sk.ca/2007-001EIS ExecutiveSummaryEnglish

[AREVA] AREVA Resources Canada Inc. 2012. McClean Lake operation technical information document, environmental performance. Vol 1.02. Saskatoon (SK): AREVA Resources Canada Inc. 832 p.

Bailey RC, Norris RH, Reynoldson TB. 2004. Bioassessment of freshwater ecosystems: Using the reference condition approach. Boston (MA): Kluwer Academic Publishers. 170 p.

[BCMOE] British Columbia Ministry of Environment. 2016. A framework for the development and use of freshwater science-based environmental benchmarks for aquatic life in environmental management act permitting for mines: Technical Guidance 8, Environmental Management Act applications, Version 1.0, March 2016. http://www2.gov.bc.ca/assets/gov/environment/waste-management/industrial-waste/industrial-waste/mining-smelt-energy/guidance-documents/tg8_framework_for_sbebs.pdf

Beanlands GE, Duinker PN. 1983. An ecological framework for studies and federal environmental assessment review office. Halifax (NS): Institute for Resource and Environmental Studies, Dalhousie University and Federal Environmental Assessment Review Office. 274 p.

Bennion H, Battarbee R, Sayer CD, Simpson GL, Davidson TA. 2011. Defining reference conditions and restoration targets for lake ecosystems using paleolimnology: A synthesis. *J Paleolimnol* 45:533–544.

Block WM, Franklin AB, Ward JP, Ganey JL, White GC. 2001. Design and implementation of monitoring studies to evaluate the success of ecological restoration on wildlife. *Restoration Ecol* 9:293–303.

Borsuk ME, Stow CA, Reckhow KH. 2002. Predicting the frequency of water quality standard violations: A probabilistic approach for TMDL development. *Environ Sci Technol* 36(10):2109–2115.

Bowman MF, Somers KM. 2005. Considerations when using the reference condition approach for bioassessment of freshwater ecosystems. *Water Qual Res J Canada* 40:347–360.

Bowman MF, Somers KM. 2006. Evaluating a novel Test Site Analysis (TSA) bioassessment approach. *J N Am Benthol Soc* 25(3):712–727.

Buckley R. 1991. Auditing the precision and accuracy of environmental impact predictions in Australia. *Environ Monit Assess* 18(1):1–23.

Burton AC, Huggard D, Bayne E, Schieck J, Solymos P, Muhly T, Farr D, Boutin S. 2014. A framework for adaptive monitoring of the cumulative effects of human footprint on biodiversity. *Environ Monit Assess* 186(6):3605–3617.

Cairns J Jr, McCormick PV, Niederlehner BR. 1993. A proposed framework for developing indicators of ecosystem health. *Hydrobiologia* 263(1):1–44.

[CCME] Canadian Council of Ministers of the Environment. 2008. Canadian water quality guidelines. Winnipeg (MB): Canadian Council of Ministers of the Environment. 1484 p. https://www.ccme.ca/files/Resources/supporting_scientific_documents/cwqg_pn_1040.pdf.

[CCME] Canadian Council of Ministers of the Environment. 2003. Guidance on the site-specific application of water quality guidelines in Canada: Procedures for deriving numerical water quality objectives. Winnipeg (MB): Canadian Council of Ministers of the Environment. 146 p. http://ceqg-rcqe.ccme.ca/download/en/221

[CCME] Canadian Council of Ministers of the Environment. 2007. A protocol for the derivation of water quality guidelines for the protection of aquatic life. Winnipeg (MB): Canadian Council of Ministers of the Environment. 37 p. http://www.ccme.ca/files/Resources/supporting_scientific_documents/protocol_aql_2007e.pdf

Chambers PA, McGoldrick DJ, Brua RB, Vis C, Culp JM, Benoy GA. 2012. Development of environmental thresholds for nitrogen and phosphorus in streams. *J Environ Qual* 41(1):7–20.

Chapman MG, Underwood AJ. 2000. The need for a practical scientific protocol to measure successful restoration. *Wetlands* 19:28–49.

[CNSC] Canadian Nuclear Safety Commission. 2014. Environmental protection framework for operating uranium mines and mills under the Nuclear Safety and Control Act. Ottawa (ON): Canadian Nuclear Safety Commission. e-DOC: 4479262. http://www.bape.gouv.qc.ca/sections/mandats/uranium-enjeux/documents/NAT26_VA.pdf

Collier KJ, Aldridge BMTA, Hicks BJ, Kelly J, Macdonald A, Smith BJ, Tonkin J. 2009. Ecological values of Hamilton urban streams (North Island, New Zealand): Constraints and opportunities for restoration. *N Z J Ecol* 33:1–13.

Cormier SM, Suter GW 2nd. 2013. A method for deriving water-quality benchmarks using field data. *Environ Toxicol Chem* 32(2):255–262.

Davies SP, Jackson SK. 2006. The biological condition gradient: A descriptive model for interpreting change in aquatic ecosystems. *Ecol Appl* 16(4):1251–1266.

DeBlois EM, Tracy E, Janes GG, Crowley RD, Wells TA, Williams UP, Paine MD, Mathieu A, Kilgour BW. 2014. Environmental effects monitoring at the Terra Nova offshore oil development (Newfoundland, Canada): Program design and overview. *Deep Sea Res Part II Top Stud Oceanogr* 110:4–12.

den Besten PJ, De Deckere E, Babut MP, Power B, DelValls TA, Zago C, Oen AMP, Heise S. 2003. Biological effects-based sediment quality in ecological risk assessment for European waters. *J Soils Sediments* 3(3): 144–162.

Dipper B, Jones C, Wood C. 1998. Monitoring and post-auditing in environmental impact assessment: A review. *J Environ Plann Man* 41(6): 731–747.

Dubé M, Munkittrick K. 2001. Integration of effects-based and stressor-based approaches into a holistic framework for cumulative effects assessment in aquatic ecosystems. *Hum Ecol Risk Assess* 7(2):247–258.

Dubé MG, Duinker P, Greig L, Carver M, Servos M, McMaster M, Noble B, Schreier H, Jackson L, Munkittrick KR. 2013. A framework for assessing cumulative effects in watersheds: An introduction to Canadian case studies. *Integr Environ Assess Manag* 9(3):363–369.

Duinker PN. 1989. Ecological effects monitoring in environmental impact assessment: What can it accomplish? *Environ Manage* 13(6):797–805.

Duinker PN, Beanlands GE. 1986. The significance of environmental impacts: An exploration of the concept. *Environ Manage* 10(1):1–10.

Environment Canada. 2003. National assessment of pulp and paper environmental effects monitoring data: A report synopsis. Burlington (ON): National Water Research Institute, Environment Canada. Series No. 2. 28 p.

Environment Canada. 2012. Metal mining environmental effects monitoring technical guidance. Gatineau (QC): Environment Canada. 544 p.

Environment Canada and Alberta Environment. 2012. Joint Canada/Alberta implementation plan for oil sands monitoring. Environment Canada. 32 p. http://www.ec.gc.ca/scitech/D0AF1423-351C-4CBC-A990-4ADA543E7181/COM1519_Final%20OS%20Plan_02.pdf

Gaudet CL, Keenleyside KA, Kent RA, Smith SL, Wong MP. 1995. How should numerical criteria be used? The Canadian approach. *Hum Ecol Risk Assess* 1(1):19–28.

Gibbons WN, Munkittrick KR. 1994. A sentinel monitoring framework for identifying fish population responses to industrial discharges. *J Aquat Ecosyst Stress Recovery* 3(3):227–237.

Glozier NE, Culp JM, Reynoldson TB, Bailey RC, Lowell RB, Trudel L. 2002. Assessing metal mine effects using benthic invertebrates for Canada's Environmental Effects Program. *Water Qual Res J Canada* 37:251–278.

Government of Alberta. 2012. Lower Athabasca regional plan, 2012–2022. 98 p. https://www.landuse.alberta.ca/LandUse%20Documents/Lower%20Athabasca%20Regional%20Plan%202012-2022%20Approved%202012-08.pdf

Green RH. 1979. Sampling design and statistical methods for environmental biologists. New York (NY): Wiley. 257 p.

Hanson N. 2011. Using biological data from field studies with multiple reference sites as a basis for environmental management: The risks for false positives and false negatives. *J Environ Manage* 92(3):610–619.

Hatfield Consultants, Kilgour & Associates Ltd, Klohn Crippen Berger, Western Resource Solutions. 2009. RAMP: Technical design and rationale. Prepared for RAMP steering committee. 348 p. http://www.ramp-alberta.org/UserFiles/File/RAMP_Design_&_ Rationale.pdf

Hatfield Consultants, Kilgour & Associates Ltd., Western Resource Solutions. 2016. Regional aquatics monitoring in support of the Joint Oil Sands Monitoring Plan, final 2015 program report. Prepared for: Alberta Environmental Monitoring, Evaluation and Reporting Agency (AEMERA). Edmonton (AB). 1272 p.

Hewitt LM. 2011. Effects-directed studies of pulp and paper mill effluents. In: Brack W, editor. Effect-directed analysis of complex environmental contamination. Handbook of environmental chemistry. No. 15. Berlin-Heidelberg (DE): Springer Verlag. p 267–283.

Hewitt LM, Dubé MG, Ribey SC, Culp JM, Lowell R, Hedley K, Kilgour B, Portt C, MacLatchy DL, Munkittrick KR. 2005. Investigation of cause in pulp and paper environmental effects monitoring. *Water Qual Res J Canada* 40:261–274.

Hewitt LM, Kovacs TG, Dubé MG, MacLatchy DL, Martel PH, McMaster ME, Paice MG, Parrott JL, van den Heuvel MR, van der Kraak GJ. 2008. Altered reproduction in fish exposed to pulp and paper mill effluents: Roles of individual compounds and mill operating conditions. *Environ Toxicol Chem* 27(3):682–697.

Holling CS. 1978. Adaptive environmental assessment and management. Chichester (UK): John Wiley & Sons. 400 p.

Huebert DB, McKernan M, Paquette C. 2011. Canada's environmental effects monitoring program: Areas for improvement. *Integr Environ Assess Manag* 7(1):143–144.

[IAEA] International Atomic Energy Agency. 2010. Best practices in environmental management of uranium mining. Report NF-T-1.2. Vienna (AT): IAEA. 47 p.

Jensen AL. 1973. Statistical analysis of biological data from preoperational-postoperational industrial water quality monitoring. *Water Res* 7(9): 1331–1347.

Kelly EN, Schindler DW, Hodson PV, Short JW, Radmanovich R, Nielsen CC. 2010. Oil sands development contributes elements toxic at low concentrations to the Athabasca River and its tributaries. *Proc Natl Acad Sci USA* 107(37):16178–16183.

Kelly EN, Short JW, Schindler DW, Hodson PV, Ma M, Kwan AK, Fortin BL. 2009. Oil sands development contributes polycyclic aromatic compounds to the Athabasca River and its tributaries. *Proc Natl Acad Sci USA* 106(52):22346–22351.

Kerr-Lawson LJ, Karrow PF, Edwards TWD, Mackie GL. 1992. A paleoenvironmental study of the molluscs from the Don Formation (Sangamonian?) Don Valley. Brickyard, Toronto, Ontario. *Can J Earth Sci* 29(11):2406–2417.

Kilgour BW, Dubé MG, Hedley K, Portt CB, Munkittrick KR. 2007. Aquatic environmental effects monitoring guidance for environmental assessment practitioners. *Environ Monit Assess* 130(1–3):423–436.

Kilgour BW, Munkittrick KR, Portt CB, Hedley K, Culp J, Dixit S, Pastershank G. 2005. Biological criteria for municipal wastewater effluent monitoring programs. *Water Qual Res J Canada* 40:374–387.

Kilgour BW, Somers KM, Barrett TJ, Munkittrick KR, Francis AP. 2017. Testing against "normal" with environmental data. *Integr Environ Assess Manag* 13(1):188–197.

Kilgour BW, Somers KM, Matthews DE. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. *Écoscience* 5(4):542–550.

Kovacs T, Hewitt M, MacLatchy D, Martel P, McMaster M, Parrott J, Van Der Kraak G, van den Heuvel M. 2007. Cycle 4 national investigation of cause final report. Gatineau (QC): National Environmental Effects Monitoring Office Quebec. 48 p.

Kovacs T, Martel P, Ricci M. 2007. A shortened adult fathead minnow reproduction test developed for investigation of cause and investigation of solution work involving pulp and paper mill effluents. *Water Qual Res J Canada* 42:91–100.

Kovacs TG, Martel PH, O'Connor BI, Parrott JL, McMaster ME, Van Der Kraak GJ, MacLatchy DL, van den Heuvel MR, Hewitt LM. 2011. Kraft mill effluent survey: Progress toward best management practices for reducing effects on fish reproduction. *Environ Toxicol Chem* 30(6): 1421–1429.

Kreutzweiser DP, Capell SS, Good KP. 2005. Macroinvertebrate community responses to selection logging in riparian and upland areas of headwater catchments in a northern hardwood forest. *J N Am Benthol Soc* 24(1):208–222.

Landrum PF. 1995. How should numerical sediment quality criteria be used? *Hum Ecol Risk Assess* 1(1):13–17.

Lowell RB, Hedley K, Porter E. 2002. Data interpretation issues for Canada's environmental effects monitoring program. *Water Qual Res J Canada* 37:101–117.

Mackey R, Rees C, Wells K, Pham S, England K. 2012. Can natural variability trigger effects on fish and fish habitat as defined in environment Canada's metal mining environmental effects monitoring program? *Integr Environ Assess Manag* 9(1):155–163.

MacLatchy DL, Milestone C, Shaughnessy KS, Belknap AM, Dubé MG, Hewitt LM. 2010. Reproductive steroid responses in fish exposed to pulp mill condensates: An investigation of cause case study. *Water Qual Res J Canada* 45:163–173.

Main C. 2011. 2010 Regional Aquatics Monitoring Program (RAMP) Scientific Review. Calgary (AB): Alberta Innovates—Technology Futures. 160 p. http://www.ramp-alberta.org/UserFiles/File/RAMP%202010%20Scientific%20Peer%20Review%20Report.pdf

Mapstone BD. 1995. Scalable decision rules for environmental impact studies: Effect size, type I, and type II errors. *Ecol Appl* 5(2):401–410.

Martel P, Hewitt M, Kovacs T, MacLatchy D, McMaster M, O'Connor B, Parrot J, van den Heuvel MR, Van Der Kraak G. 2010. Cycle 5 national investigation of cause project final report. Gatineau (QC): National Environmental Effects Monitoring Office. 95 p.

McBride GB, Loftis JC, Adkins NC. 1993. What do significance tests really tell us about the environment? *Environ Manage* 17(4):423–432.

McMaster ME, Frank M, Munkittrick K, Riffon R, Wood C. 2002. Follow-up studies addressing questions identified during cycle 1 of the adult fish survey of the pulp and paper EEM program. *Water Qual Res J Canada* 37:133–153.

Mellina E, Hinch SG. 1995. Overview of large-scale ecological experimental designs and recommendations for the British Columbia Watershed Restoration Program. Watershed Restoration Project Report 1. British Columbia (CA): Ministry of Environment, Lands and Parks, and Ministry of Forests. 31 p.

Miskimmin B, Fedorak P, Lauman R, Vinke K. 2010. Oil sands water toxicity: A critical review. Report 2010-089. Devon (AB): Natural Resources Canada and Canmet ENERGY. 106 p.

Munkittrick KR, Dixon DG. 1989a. An holistic approach to ecosystem health assessment using fish population characteristics. *Hydrobiologia* 188–189(1):123–135.

Munkittrick KR, Dixon DG. 1989b. Use of white sucker (*Catostomus commersoni*) populations to assess the health of aquatic ecosystems exposed to low-level contaminant stress. *Can J Fish Aquat Sci* 46(8): 1455–1462.

Munkittrick KR, Barrett TJ, McMaster ME. 2010. Guidance for site-specifically assessing the health of fish populations with emphasis on Canada's environmental effects monitoring program. *Water Qual Res J Canada* 45:209–221.

Murphy RB. 1948. Non-parametric tolerance limits. *Ann Math Stat* 19(4):581–589.

Neff JM, Bean DJ, Cornaby BW, Vaga RM, Gulbransen RC, Scanlon JA. 1986. Sediment quality criteria methodology validation: Calculation of screening level concentrations from field data. Battelle (WA): Environmental Program Office for USEPA. 243 p.

Nemec AFL. 1998. Design of experiments. In: Sit V, Taylor B, editors. Statistical methods for adaptive management studies. Land management handbook. No. 42. Victoria (BC): Ministry of Forests and Range Research Branch. p 9–18.

Oil Sands Advisory Panel. 2010. A foundation for the future: building an environmental monitoring system for the oil sands. A report submitted to the federal Minister of Environment by the Oil Sands Advisory Panel for the Lower Athabasca River and connected waterways. Government of Canada. 49 p. http://publications.gc.ca/collections/collection_2011/ec/En4-148-2010-eng.pdf.

[OMOEE] Ontario Ministry of Environment and Energy. 1993. Guidelines for the protection and management of aquatic sediment in Ontario. Etobicoke (ON): Queen's Printer for Ontario. 39 p.

[OMOEE] Ontario Ministry of Environment and Energy. 1994. Water management, policies, guidelines, provincial water quality objectives of the Ministry of Environment and Energy. Etobicoke (ON): Queen's Printer for Ontario.

Osenberg CW, Schmitt RJ, Holbrook SJ, Abu-Saba KE, Flegal AR. 1994. Detection of environmental impacts: Natural variability, effect size, and power analysis. *Ecol Appl* 4(1):16–30.

Paine MD, DeBlois EM, Kilgour BW, Tracy E, Pocklington P, Crowley RD, Williams UP, Gregory Janes G. 2014. Effects of the Terra Nova offshore oil development on benthic macro-invertebrates over 10 years of development drilling on the Grand Banks of Newfoundland, Canada. *Deep Sea Res Part II Top Stud Oceanogr* 110:38–64.

Parsons BG, Watmough SA, Dillon PJ, Somers KM. 2010. A bioassessment of lakes in the Athabasca Oil Sands Region, Alberta, using benthic macroinvertebrates. *J Limnol* 69:105–117.

Pearson TH, Rosenberg R. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanogr Mar Biol* 16:229–311.

Quinn G, Keogh M. 2002. Experimental design and data analysis for biologists. Cambridge (UK): Cambridge University Press. 557 p.

RAMP Technical Program Committee and Implementation Team. 2011. Scientific peer review of RAMP: Response to panel comments and recommendations. Final report to the RAMP scientific peer review panel. Vancouver (BC): Hatfield Consultants. 167 p. http://www.ramp-alberta.org/UserFiles/File/Supporting/FinalResponsetoRAMPPeerReview.pdf

Ribey SC, Munkittrick KR, McMaster ME, Courtenay S, Langlois C, Munger S, Rosaasen A, Whitley G. 2002. Development of a monitoring design for examining effects in wild fish associated with discharges from metal mines. *Water Qual Res J Canada* 37:229–249.

Richter BD, Baumgartner JV, Powell J, Braun DP. 1996. A method for assessing hydrologic alteration within ecosystems. *Conserv Biol* 10(4):1163–1174.

Roux DJ, Kempster PL, Kleynhans CJ, van Vliet HR, Du Preez HH. 1999. Integrating stressor and response monitoring into a resource-based water-quality assessment framework. *Environ Manage* 23(1):15–30.

Scarsbrook MR, Boothroyd IKG, Quinn JM. 2000. New Zealand's National River Water Quality Network: Long-term trends in macroinvertebrate communities. *N Z J Mar Freshw Res* 34(2):289–302.

Schindler D. 2010. Tar sands need solid science. *Nature* 468(7323):499–502.

Schwarz CJ. 1998. Studies of uncontrolled events. In: Sit V, Taylor B, editors. Statistical methods for adaptive management studies. Land management handbook. No. 42. Victoria (BC): Ministry of Forests and Range Research Branch. p 19–39.

Scott WB, Crossman EJ. 1973. Freshwater fishes of Canada. Bulletin 184. Ottawa (ON): Fisheries Research Board of Canada. 966 p.

Smith MPL. 1991. Environmental impact assessment: The roles of predicting and monitoring the extent of impacts. *Aust J Mar Freshwater Res* 42(5): 603–614.

Smith RA, Alexander RB, Schwarz GE. 2003. Natural background concentrations of nutrients in streams and rivers of the conterminous United States. *Environ Sci Technol* 37(14):3039–3047.

Smol JP, Stoermer EF. 2010. The diatoms: Applications for the environmental and earth sciences. 2nd ed. Cambridge (UK): Cambridge University Press. 686 p.

Spellerberg I. 2005. Monitoring ecological change. 2nd ed. Cambridge (UK): Cambridge University Press.

Stewart-Oaten A, Bence JR, Osenberg CW. 1992. Assessing effects of unreplicated perturbations: No simple solutions. *Ecology* 73(4):1396–1404.

Stewart-Oaten A, Murdoch WM, Parker KR. 1986. Environmental impact assessment: 'Pseudoreplication' in time? *Ecology* 67(4):929–940.

Suter G. 1993. Ecological risk assessment. Boca Raton (FL): Lewis Publishers.

Thompson PA, Kurias J, Mihok S. 2005. Derivation and use of sediment quality guidelines for ecological risk assessment of metals and radionuclides released to the environment from uranium mining and milling activities in Canada. *Environ Monit Assess* 110(1–3):71–85.

Underwood AJ. 1991. Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Aust J Mar Freshwater Res* 42(5):569–587.

Underwood AJ. 1992. Beyond BACI: The detection of environmental impacts on populations in the real, but variable world. *J Exp Mar Biol Ecol* 161(2):145–178.

Underwood AJ. 1994. On beyond BACI: Sampling designs that might reliably detect environmental disturbances. *Ecol Appl* 4(1):3–15.

Underwood AJ, Chapman MG. 2003. Power, precaution, type II error and sampling design in assessment of environmental impacts. *J Exp Mar Biol Ecol* 296(1):49–70.

[USEPA] United States Environmental Protection Agency. 1986. Quality criteria for water, 1986. EPA 440/5-86-001. Washington (DC): United States Environmental Protection Agency, Office of Water, Regulations and Standards. 395 p.

Walker IR. 1987. Chironomidae (Diptera) in paleoecology. *Quat Sci Rev* 6(1): 29–40.

Walker SL, Hedley K, Porter E. 2002. Pulp and paper environmental effects monitoring in Canada: An overview. *Water Qual Res J Canada* 37:7–19.

Walker SL, Lowell RB, Sherry JP. 2005. Preliminary analysis of pulp and paper environmental effects monitoring data to assess possible relationships between the sublethal toxicity of effluent and effects on biota in the field. *Water Qual Res J Canada* 40:251–260.

Walker SL, Ribey SC, Trudel L, Porter E. 2003. Canadian environmental effects monitoring: Experiences with pulp and paper and metal mining regulatory programs. *Environ Monit Assess* 88(1–3):311–326.

Westgard JO, Barry PL, Hunt MR, Groth T. 1981. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clin Chem* 27(3):493–501.

White MS, Xenopoulos MA, Metcalfe RA, Somers KM. 2011. Water level thresholds of benthic macroinvertebrate richness, structure, and function of boreal lake stony littoral habitats. *Can J Fish Aquat Sci* 68(10): 1695–1704.

Wissmar RC. 1993. The need for long-term stream monitoring programs in forest ecosystems of the Pacific Northwest. *Environ Monit Assess* 26(2–3):219–234.

Wood C, Dipper B, Jones C. 2000. Auditing the assessment of the environmental impacts of planning projects. *J Environ Plann Manage* 43(1):23–47.

Zajdlik BA. 2016. A statistical evaluation of the safety factor and species sensitivity distribution approaches to deriving environmental quality guidelines. *Integr Environ Assess Manag* 12(2):380–387.

